Monitoring, Diagnostics and Accounting

TANYA LEVSHINA

Talk Overview

- New Infrastructure
- Offsite Submission
- Monitoring and Debugging:
  - Grid Jobs
  - Data Management
- Accounting
- Help!

What's New and Why Should I Care? (I)

- New jobsub: client/server architecture
  - More scalable than old architecture
  - Changes that affect you:
    - You cannot login on submission node
    - jobsub commands have changed – check help & documentation
    - Log files need to be fetched from the server site
    - jobsub will show the status of your jobs but if you want to see job details for now you should use the command condor_q (and specify pool and schedd)
- High Availability Service
  - Multiple jobsub servers
  - Changes that affect you:
    - new job id:cluster.process@jobsubserver

What's New and Why Should I Care? (II)

- New software distribution paradigm: OASIS/CVMFS
  - Robust, scalable way to distribute software on remote sites
  - Changes that affect you:
    - You need to build relocatable software
    - Higher probability of encountering access problems on remote worker nodes
- New resources: FermiCloud, the OSG(more than 80 institutions), and AWS clusters:
  - More opportunistic resources
  - Changes that affect you:
    - Complicated troubleshooting
    - Cannot expect to have access to BlueArc

Gird Jobs: Monitoring and Debugging

Job Submission

You submitted a job:
<> jobsub_submit.py --resource-provides=usage_model=DEDICATED --group uboone ....
Submitting job(s)..........
10 job(s) submitted to cluster 386.
JobsubJobId of first job: 386.0@fifebatch2.fnal.gov
Use job id 386.0@fifebatch2.fnal.gov to retrieve output

Let's look at the status job (386.0@fifebatch2.fnal.gov) :
<> jobsub_q.py -G uboone

386.0@fifebatch2.fnal.gov	tlevshin	06/10	15:39	0+00:00:00	I
0	0.0	test_local_env.sh			
386.1@fifebatch2.fnal.gov	tlevshin	06/10	15:39	0+00:00:00	I
0	0.0	test_local_env.sh			
386.2@fifebatch2.fnal.gov	tlevshin	06/10	15:39	0+00:00:00	I
0	0.0	test_local_env.sh			

Why my job is idle?

Idle jobs (I)

There are multiple reasons why your job can be idle:

- It requires some time to request and start a glidein on a remote worker node and match your job. Be patient especially if your are submitting jobs to opportunistic, offsite, fermicloud or AWS resources.

- Your job doesn't match any requirements
  - Unsupported resources (SL7, huge memory requirement)

- Your job is starting, failing right away and it is put back into a queue

- The system is too busy running other jobs.

Idle jobs (II)

If your job is idle for more than half an hour, check:

- requirements using condor_q –better-analyze.
- for number of restarts:

  condor_q -g -pool fifebatch1.fnal.gov –l <job_id> -format "%s\n" NumRestarts

  NumRestarts = 5

- the log file and look for memory usage:

  ....

  006 (528.001.000) 06/12 13:33:42 Image size of job updated: 14636

      2 - MemoryUsage of job (MB)

      1668 - ResidentSetSize of job (KB)
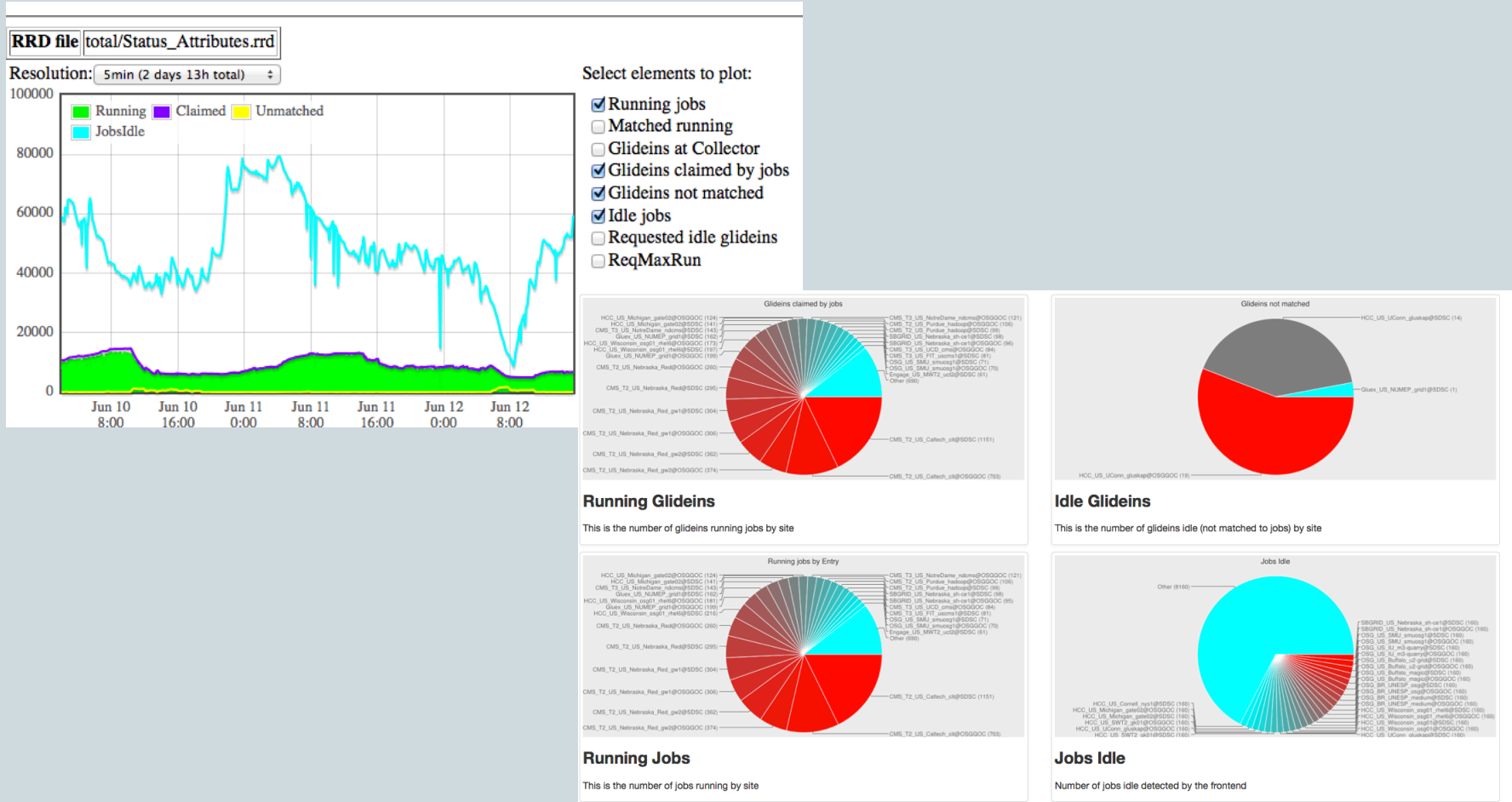
  ...

- GlideinWMS Frontend monitoring page.

  http://fifebatchgpvmhead1.fnal.gov/vofrontend/monitor

VO Frontend Monitor (http://fifebatchgpvmhead1.fnal.gov/vofrontend/monitor)

Held Jobs

If you see:

condor_q -g -pool fifebatchgpvmhead1.fnal.gov -name fifebatch2.fnal.gov
ID   OWNER   SUBMITTED   RUN_TIME ST PRI SIZE CMD
7741.0@fifebatch2.fnal.gov   tlevshin   6/5 10:20   0+00:00:28 H 0   0.0 test.sh

Check the held reason:

condor_q -g -pool fifebatchgpvmhead1.fnal.gov –name fifebatch2.fnal.gov
  –l 7741.0@fifebatch2.fnal.gov
  -format "%s " MATCH_EXP_JOB_GLIDEIN_Site -format "%s\n" HoldReason

MATCH_EXP_JOB_GLIDEIN_Site = "BNL"

HoldReason = "Error from glidein_7363@acas1505.usatlas.bnl.gov: error changing sandbox ownership to the user: condor_glexec_setup exited with status 768 Glexec exited with status 203;"

Other reasons:

HoldReason = "Job restarted too many times (by user condor)"

HoldReason = "Error from glidein_24733@compute-1-2.nys1: STARTER at 192.168.9.253 failed to send file(s) to <129.79.53.21:9615>: error reading from …SHADOW failed to receive file(s) from <128.84.3.204:37410>"

Open SNOW ticket!

Job Monitoring

- Use jobsub_q.py to check if your jobs is still running
- Use jobsub_fetchlog to download your log files
- Check FIFEMON (we plan to support FIFEMON for the new infrastructure)

  http://fifemon1.fnal.gov/monitor/user/<username> - *user's jobs info*

  http://fifemon1.fnal.gov/monitor/experiment/<vo> - *batch details, enstore, blue arc and dCache activities per experiment*

Batch details about your experiment

Storage details

Batch details about you job

Job ID	Group	Submit time	Procs	Running	Queued	Held	Completed	Restarts	Cluster	Command
7741		Jun 05 10:20 AM	10	0	0	10		0	fifebatch1	/fife/local/data/rexbatch/uploads/lbne/tlevshin /2014-06-05_102028.939960_1448/test_local_env.sh

Job Debugging (I)

Job failed:

- Do you have enough information to debug a problem
  - Add set –x for simplify debugging
  - *Exit with non-zero exit code in case of failure*
  - *Add timestamps to your log (echo " Started ifdh `date `") especially when you are doing file transfers.*
  - You should always log the following debugging information: site, host, and OS your job has been executed:

    echo Site:${GLIDEIN_ResourceName}

    echo "the worker node is " `hostname` "OS: " `uname –a `

- Your script has a bug. Have you tested it on your local machine?
- Authentication/authorization problems with remote site

Job Debugging (II)

- Worker node problems:
  - cvmfs problem
  - missing libraries
  - not enough space (worker node scratch area per slot is limited to 20GB – 80GB)
  - not enough memory (OSG limit is usually 2GB per slot)

Create a SNOW ticket, provide:

- Your experiment, role (DN and FQAN of a proxy certificate)
- On what Site, Node your job was executed
- Job stdout/stderr and log files

You can always request a FermiCloud VM node that looks like a "generic" grid worker node and do testing and debugging there.

# Data Management: Monitoring and Debugging

Data Management

- We're not in Kansas anymore! The jobs could be started at Nebraska, Oklahoma or other OSG sites. So you should not expect to get direct access to BlueArc.
  - How do you get files to/from a worker node?
- Do you use SAM/SAM WEB? Check your SAM station for input file download:

  http://samweb.fnal.gov:8480/station_monitor/<vo>

- Do you use IFDH? Do you know where your output is going: Fermi Public dCache? Fermilab BlueArc? Remote Storage Element?

SAM Monitoring

Project Id	16666
Status	running
Owner	satish
Start time	2014-06-11 08:03:29
Dataset definition	prod_fardet_artdaq_cosmics_15400_16000_draining_S13-03-24
Files in snapshot	3543
Files seen	3543
Processes	179
Busy processes	0
Finished processes	178
Waiting processes	0
Error processes	1
Mean wait time (per file)	1s
Mean busy time (per file)	21min 45s
Last activity	process ended at 2014-06-11 16:33:36
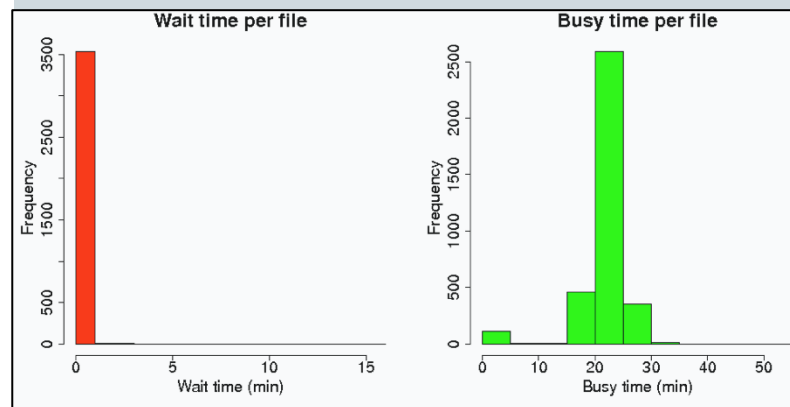
Processes

Process Id	Node name	Status	Description	Files seen	Last change	Waiting for	Mean wait time (per file)	Mean busy time (per file)
732712	fnpc3002.fnal.gov	error	16625758.8	1	2014-06-11 08:21:29 (process ended - error)	-	7min 40s	-
732713	fnpc5015.fnal.gov	completed	16625758.0	20	2014-06-11 14:32:36 (process ended - completed)	-	0s	19min 24s
732714	fnpc4005.fnal.gov	completed	16625758.4	20	2014-06-11 15:17:43 (process ended - completed)	-	0s	21min 39s

IFDH Monitoring

- It is strongly recommended to add to your script or pass it with –e to jobsub (especially for jobs running offsite):

  export IFDH_DEBUG=1

- ifdh does a lot of guessing how to reach a proper storage when plain path is specified
  https://cdcvs.fnal.gov/redmine/projects/ifdhc/wiki/Smarter_data_routing

- If you need to use the remote storage element, you will need to know the address and path to the storage element and use it as the source or destination arguments in the ifdh command.

IFDH Debugging (I)

- What can go wrong (especially when running offsite and/or using remote SEs)?:
  - Service (SRM Endpoint or gridftp) is down
  - Network problems
  - Authentication (proxy has expired)
  - Authorization (you don't have access to a SE or permissions to access a specific directory)
  - Worker node problem:
    - missing executables, libraries (globus, srm client commands) or cvmfs problems
    - problem with a host certificate or crl files
  - File not found, wrong path , etc – trivial bugs

IFDH Debugging (II)

- To debug:
  - Check logs
  - Execute ifdh command manually
- Create a SNOW ticket, provide:
  - Your experiment, role (DN and FQAN of a proxy certificate)
  - On what Site, Node your job was executed
  - What command you have used to transfer files
  - Timestamp
  - Command stdout/stderr output
  - Results of executing the same command manually in local environments

dCache Monitoring

You can check dCache transfers if your files are going to Fermilab Public dCache (path /pnfs/fnal.gov/usr/<vo>/ scratch/…)

- Home page: http://fndca.fnal.gov

- Per experiment space usage: http://fndca.fnal.gov/cgi-bin/sg_usage_cgi.py

- Per experiment dCache transfers: http://fndca.fnal.gov/cgi-bin/sg_transfers_cgi.py

- Active transfers: http://fndca3a.fnal.gov:2288/context/transfers.html

- Recent FTP transfers: http://fndca.fnal.gov/cgi-bin/dcache_files.py

- File lifetime plots: http://fndca.fnal.gov/dcache/dc_lifetime_plots.html

novaana

Time	User	Type	Oper	File	Node	dT 1	File Size	dT 2	Status	Details
2014-06-12 14:00:16	novaana	gsiftp	read	/pnfs/fnal.gov/usr/nova/production/raw2root/S14-03-25/fardet/000149/14965/02/fardet_r00014965_s14_t02_S14-03-25_v1_data.artdaq.root	novagpvm01.fnal.gov	2	216020901	0	OK	
2014-06-12 13:59:11	novaana	gsiftp	read	/pnfs/fnal.gov/usr/nova/production/raw2root/S14-03-25/fardet/000149/14963/02/fardet_r00014963_s03_t02_S14-03-25_v1_data.artdaq.root	novagpvm01.fnal.gov	1	213602311	0	OK	
2014-06-12 13:58:04	novaana	gsiftp	read	/pnfs/fnal.gov/usr/nova/production/raw2root/S14-03-25/fardet/000149/14989/02/fardet_r00014989_s52_t02_S14-03-25_v1_data.artdaq.root	novagpvm01.fnal.gov	2	213538663	0	OK	

Space usage on FNDCA by Storage group

Data generated on 2014-06-12 15:05:44 (updated hourly) [1]

Disk space usage by Storage Group, PoolGroup (GiB)

Storage Group	ArchivePools	DESPools	EmptyPools	ExpDbWritePools	FermigridVolPools	MinervaWritePools	NovaWritePools	PublicScratchPools	RawDataWritePools	TDESPools	readWritePools	Total
nova	0	0	0	0	0	0	38,998	8,015	0	0	497,030	544,04
lqcd	0	0	0	0	0	0	0	0	0	0	410,646	410,64
minerva	0	0	0	0	0	6,129	0	7,679	0	0	326,585	340,39
darkside	0	0	0	0	0	0	0	0	0	0	332,188	332,18
none	0	0	0	0	0	0	0	0	0	0	104,812	104,81
minos	0	0	0	0	0	0	0	0	13,074	0	73,246	86,32

Active Transfers

Door	Domain	Seq	Prot	Owner	Proc	PnfsId	Pool	Host	Status	Since	S	Trans. (KB)	Speed (KB/s)
DCap01-fndca4a--602713	dcap01-fndca4aDomain	3	dcap-3	11469	4682	000002B484D600104259896C08540AFD33F1	rw-stkendca19a-1	ds50samgpvm01.fnal.gov	WaitingForDoorTransferOk	00:11:42	RUNNING	184139439	262123

Enstore Monitoring

You can check Enstore (path /pnfs/fnal.gov/usr/<vo>/rawdata)

* Home page: http://www-stken.fnal.gov/enstore/

* Small Files Aggregation: http://www-stken.fnal.gov/cgi-bin/enstore_sfa_hud_cgi.py

* Per experiment tape burn-rate: http://www-stken.fnal.gov/enstore/all_sg_burn_rates.html

* Quotas (more usage): http://www-stken.fnal.gov/enstore/tape_inventory/VOLUME_QUOTAS

* Per experiment drive usage: http://www-stken.fnal.gov/enstore/drive-hours-sep/plot_enstore_system.html

* Complete file listings: http://www-stken.fnal.gov/cgi-bin/enstore_file_listing_cgi.py

Enstore Monitoring Examples

Month	2014-01		2014-02		2014-03		2014-04		2014-05		2014-06	
Storage Group	**Store**	**Restore**	**Store**	**Restore**	**Store**	**Restore**	**Store**	**Restore**	**Store**	**Restore**	**Store**	**Restore**
HCC	13	0	2,419	0	2,420	1	0	0	0	0	0	0
astro	944	0	0	372	0	0	0	0	0	0	0	0
cdms	4,260	0	2,363	0	2,366	0	1,976	270	2,706	2	915	0
coupp	0	0	73	0	0	0	139	174	0	0	0	0
damic	0	0	0	0	38	0	246	0	265	0	39	0

Store/Restore to/from Enstore by Storage Groups. Last 6 Months (GiB)
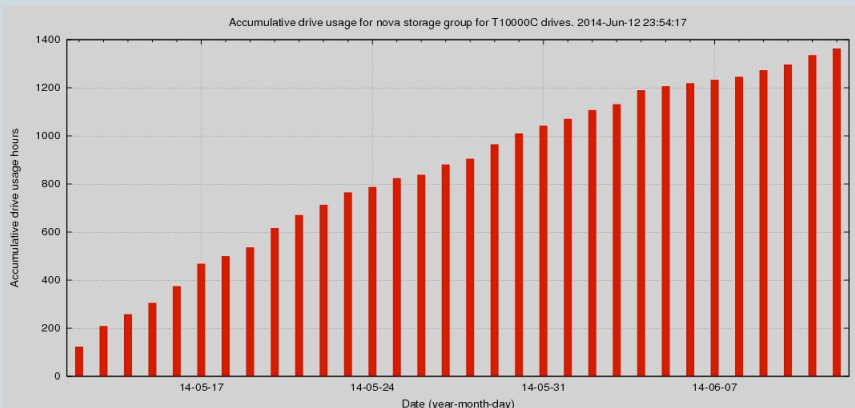
Drive usage per experiment
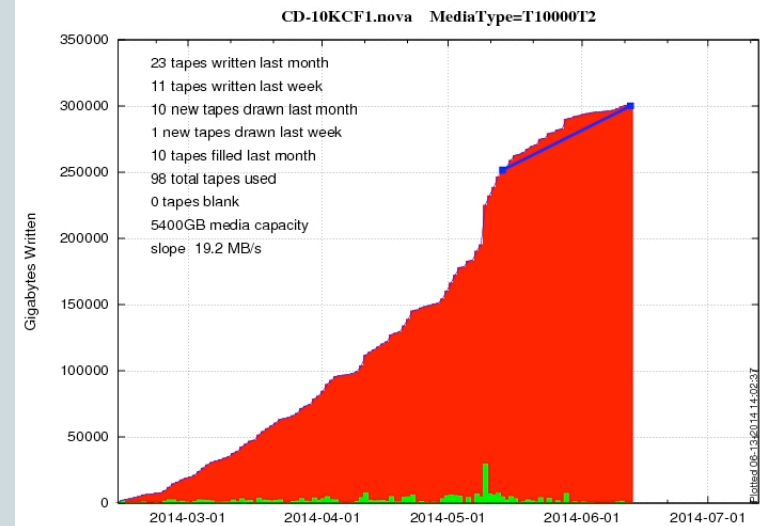
Tape burn rate per experiment

FTS Monitoring

- The File Transfer Service (FTS) is used for adding new files into the SAM catalogue, copying them to new locations, and archiving them to Enstore.

- Example: http://novadaq-ctrl-datadisk-02.fnal.gov:8888/fts/status

FTS status for novadaq-ctrl-datadisk-02.fnal.gov

Generated at 2014-06-13 15:54:08 CDT (refresh)
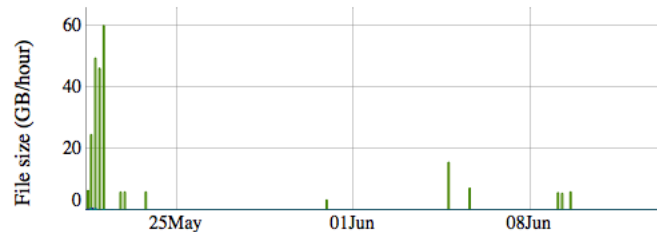
Summary

FTS: OK	FSS: OK	Stager: OK

Completed files:	147395
Failed transfers:	2276
All error files:	34553
Pending files:	580
New files:	436

Data transferred

Transfer rate
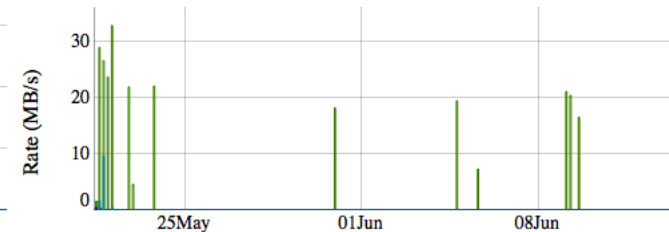
- Recent completed transfers

Time	File name	Destination
raw		
2014-06-12 11:01:06 CDT	ndos_r00016376_s15_ddtSlowMo.raw	enstore:/pnfs/nova/rawdata/NDOS/runs/000163/16376
merged-log		
2014-06-09 15:06:08 CDT	log_merge_ids_7442761_7447110_1402332927.tar	enstore:/pnfs/nova/archived_logs/2014/06

Accounting

Grid Accounting (Gratia)

- Fermi Accounting scoreboard:

  http://web1.fnal.gov/scoreboard/

  ○ Weekly generated customizable plots (wall duration, job count, VMs usage, transfers to dCache and BlueArc, usage of the OSG).

  ○ Weekly and monthly generated customizable plots per experiments

  ○ We can add more plots on request.
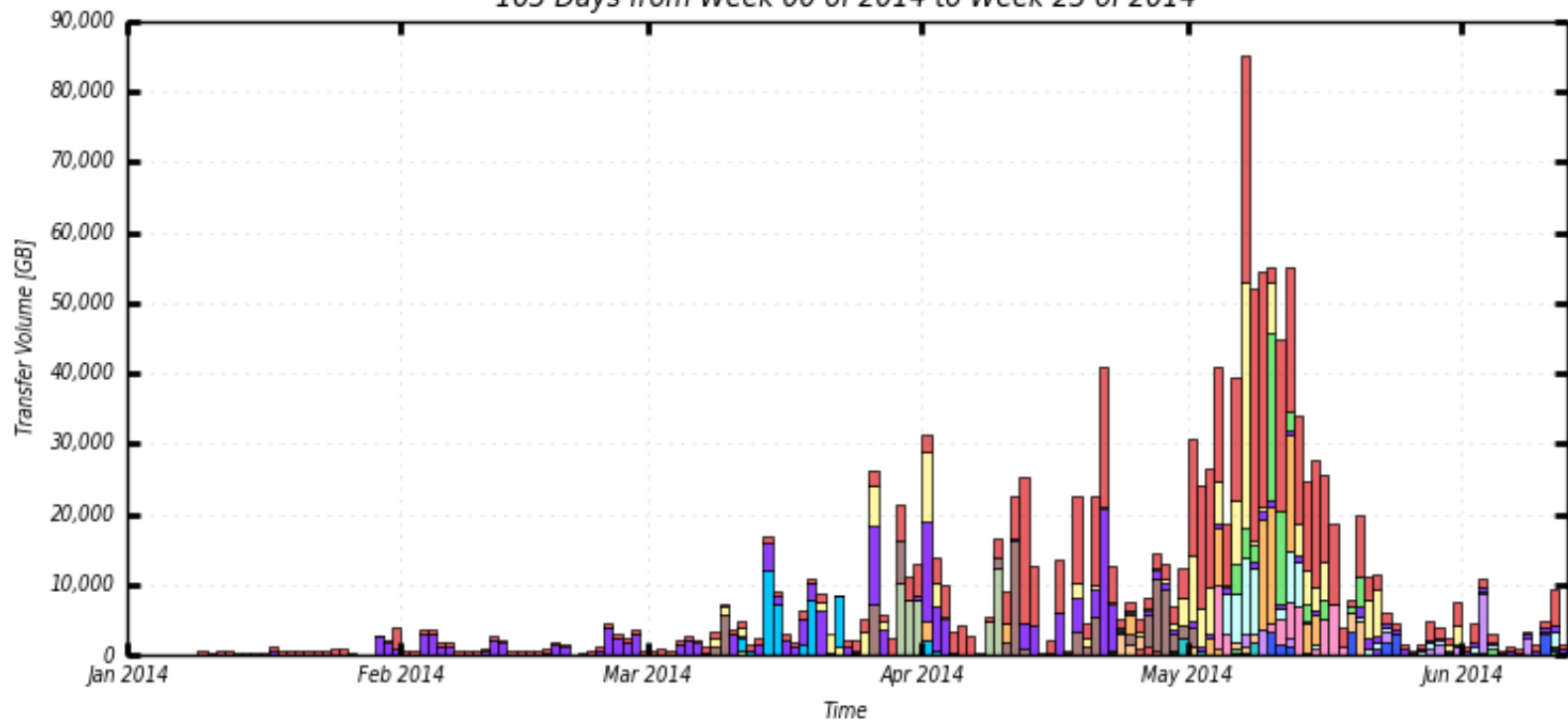
Accounting Examples(I)

Experiment Accounting:
http://web1.fnal.gov/
scoreboard/<exp>_week.html

Accounting Examples (II)

Volume of Gigabytes Transferred By User
163 Days from Week 00 of 2014 to Week 23 of 2014

Legend:
- FTS
- RUTH B. TONER
- SATISH DESAI
- BARNALI CHOWDHURY
- PATRICK T. LUKENS
- DOMINICK ROCCO
- GAVIN S. DAVIES
- NATHAN S. MAYER
- JAN ZIRNSTEIN
- NICHOLAS RADDATZ
- ERIC FLUMERFELT
- MATTHEW C. TAMSETT
- NITIN YADAV
- JIANMING BIAN
- TIAN XIN
- ADAM J. AURISANO
- CHRISTOPHER J. BACKHOUSE
- EVAN D. NINER
- SIVA KASETTI
- Other

Maximum: 85,303 GB, Minimum: 0.00 GB, Average: 8,521 GB, Current: 1,461 GB

Efficient usage of the OSG sites (CPU wall hours/Wall hours):

Wall Efficiency Quality Map by Site for nova
7 Days from 2014-06-02 00:00 to 2014-06-08 23:59

Wall Efficiency Quality Map by Site for nova
32 Days from 2014-05-08 to 2014-06-08
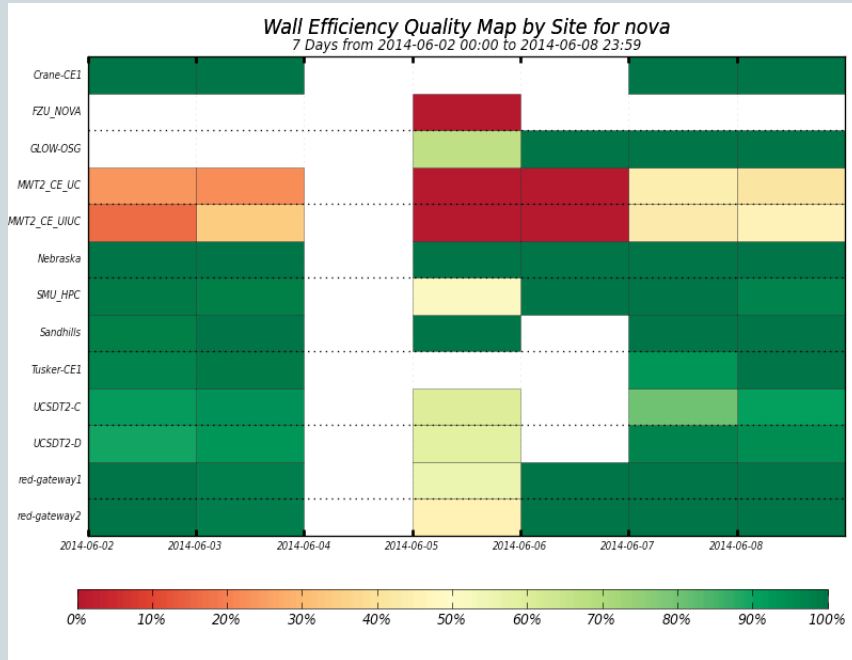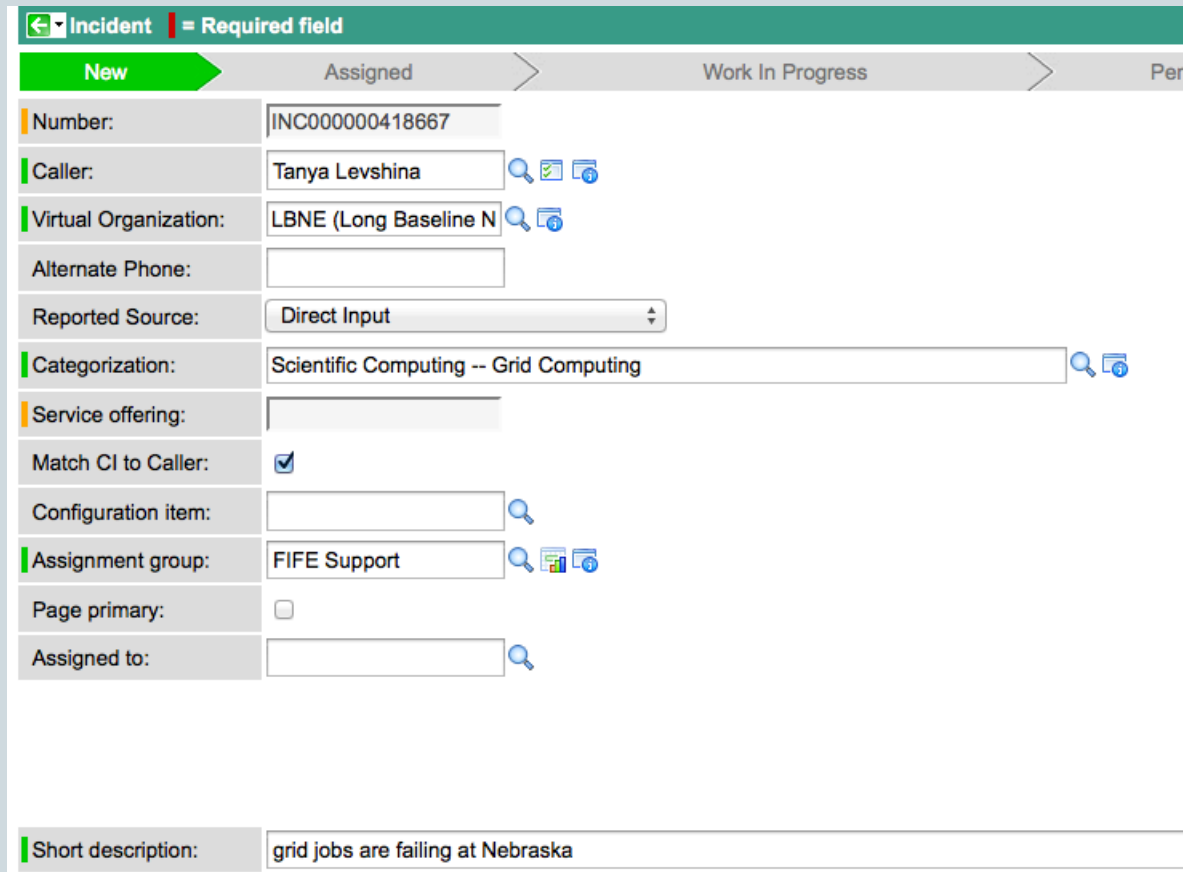
It looks like MWT2 sites were not very reliable sites for NOvA during the last week. Let's check for a monthly trend.

Getting Help

Open a SNOW ticket: https://fermi.service-now.com/